1   Overview

On Monday we studied cascades, and studied how an adopted behavior may propagate throughout the entire network (a complete cascade) given the right initial set of adopters. As we saw, the tendency of a behavior to spread in a network depends on the presence of clusters of nodes that may prevent behavior adoption. In this lecture, we examine a similar model for epidemics to answer the following questions:

- How does a disease spread?

- What's the probability that a disease stops spreading?

- Alternatively, what's the probability that a disease continuously spreads (potentially forever, given enough targets)?

2   A Simple Model of Contagion

Let us begin by defining our model formally.

> **Definition.** *Our contagion model is based around people (nodes) and phases (the levels in the network), as follows:*
>
> - *at phase $i = 0$, there is exactly one infected person;*
>
> - *at phase $i \geq 1$, each person from phase $i - 1$ meets k new, distinct individuals, and, if they were infected in phase $i - 1$, infects each new person with probability p;*
>
> - *if a person in the previous phase infects a person in the next phase, we draw a directed edge from the infecting node in the previous phase to the infected node in the next phase*

Figures 1 shows an example of how we might represent such a model as a network. Note that in this model we assume our population is of infinite size, so a disease cannot stop spreading simply because there are no individuals left to infect. In such a network, we are interested in determining the probability that a disease stops spreading (does not get passed on to anyone after a certain phase), as well as the probability that the disease never dies out (it gets endlessly passed on to future phases). To answer these questions, we give the following definition:

> **Definition.** *A contagion model with parameters p, k, has* **reproductive number** $R_0 = pk$.
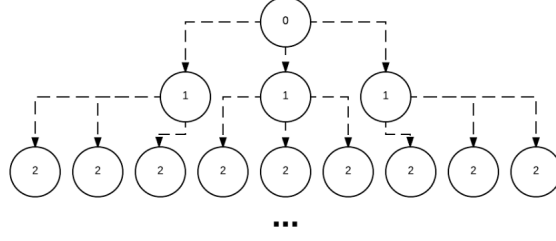
Figure 1: The first 3 phases of a simple contagion model with $k = 3$. The number in each node indicates the phase that node is in. The dashed directed edges are a placeholder to indicate that any such edge may exist with probability p if the source node of that edge is infected.

Given this definition we now propose the following theorem:

Theorem. *For a contagion model network with reproductive number R_0:*

1. *If $R_0 < 1$, then the disease will eventually stop spreading with probability 1.*

2. *If $R_0 > 1$, then the disease will never die out with probability greater than 0.*

Proof. Before we begin our proof to this theorem, we propose the following definitions:

$$q_n = Pr(\exists \text{ at least one person infected at phase } n)$$

$$q^* = \lim_{n \to \infty} q_n$$

We then have the following claims:

1. When $R_0 < 1$, $q^* = 0$

2. When $R_0 > 1$, $q^* > 0$

Note that statements 1 and 2 from the theorem are equivalent to claims 1 and 2, respectively, and so proving each of these claims is equivalent to proving the theorem.

- *Proof of Claim 1* At phase n, there are k^n nodes in that level of the network. Also, the probability of a node j at phase n of being infected is p^n. Let X_n be a random variable denoting the number of nodes infected at phase n, and let Y_{nj} be a random variable with value 1 if node j at phase n is infected, and 0 otherwise. We thus have

$$X_n = \sum_{j=1}^{k^n} Y_{nj}$$

$$E(X_n) = \sum_{j=1}^{k^n} E(Y_{nj})$$

$$= \sum_{j=1}^{k^n} p^n$$

$$= (kp)^n$$

$$= (R_0)^n$$

2

Working from the standard definition of expected value, we also have

$$E(X_n) = \sum_{i=1}^{\infty} i * Pr(X_n = i)$$

$$= \sum_{i=1}^{\infty} Pr(X_n \geq i)$$

$$\geq Pr(X_n \geq 1)$$

$$= q_n$$

Since q_n is a probability, we know that it has a lower bound value of zero, and we have just shown that it has an upper bound value of $E(X_n)$. Given that $R_0 < 1$ in this case, we combine our findings to get the following:

$$0 \leq \lim_{n \to \infty} q_n$$

$$= q^*$$

$$\leq \lim_{n \to \infty} E(X_n)$$

$$= \lim_{n \to \infty} (R_0)^n$$

$$= 0$$

Therefore, if $R_0 < 1$, we have shown that $q^* = 0$, completing the proof of Claim 1. QED

- *Proof of Claim 2* Consider a node j in phase 1 of our contagion model with parameters p, k. The probability that there exists at least one infected person at phase n that was infected from a branch through j is then pq_{n-1}. Then the probability that such a person does *not* exist is $1 - pq_{n-1}$. Given that there are k such nodes j in phase 1, the total probability that there are no infected people at phase n is then $(1 - pq_{n-1})^k$. Thus, the probability that there exists at least one infected person at phase n, or q_n, is

$$q_n = 1 - (1 - pq_{n-1})^k$$

We now have a recursive formula for q_n. To determine the properties of q_n, we can write the recursive formula instead as a function $f(x) = 1 - (1 - px)^k$. If we denote $f_{n+1}(x) = f(f_n(x))$ for $n \geq 2$ with $f_1(x) = f(x)$, then given that $q_0 = 1$, we have $q_1 = f_1(x)$, $q_2 = f(f_1(x)) = f_2(x)$, $q_3 = f(f_2(x)) = f_3(x)$, and so on. Therefore, to prove that $q^* = \lim_{n \to \infty} q_n > 0$, we can equivalently show that $\lim_{n \to \infty} f_n(1) > 0$.

To prove that $\lim_{n \to \infty} f_n(1) > 0$, we first determine some properties of $f(x)$:

1. $f(0) = 0$, and $f(1) = 1 - (1 - p)^k < 1$

2. Taking the derivative with respect to x, we have $f'(x) = pk(1 - px)^k$. For the domain $x = [0, 1]$, $f'(x)$ is a positive, decreasing function, leading us to conclude that $f(x)$ is a concave down function from $x = 0$ to $x = 1$.
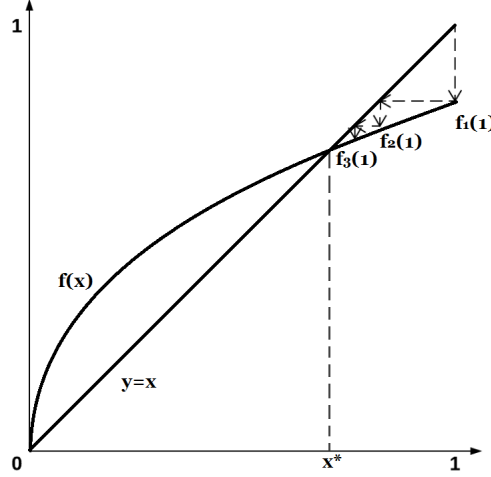
3. $f'(0) = pk = R_0 > 1$

Figure 2: Plot of f(x) and y=x, visually demonstrating our proof that $\lim_{n\to\infty} f_n(1) = x^* > 0$

Given these three properties, we plot the functions $f(x)$ and $y = x$ in Figure 2, restricting our domain to $x = [0, 1]$ since x represents a probability. As is visually observed in the plot, since $f'(0) > 1$ and $f(0) = 1$, for some small positive values of x our $f(x)$ curve will be above $y = x$. However, since $f(1) < 1$, we know that at $x = 1$ our $f(x)$ curve will be below $y = x$, producing the resulting concave plot for $f(x)$ that we see in Figure 2. We notice that if we start at $f_1(1)$ on the $f(x)$ curve, if we repeatedly take the function of $f_1(1)$ the resulting value ($f_n(1)$ with increasing n) will gradually approach the point at which the two functions $f(x)$ and $y = x$ intersect, namely for some value x^*. Since $f(x)$ is initially greater than x and eventually becomes less than x within the range [0,1], we know that $x^* > 0$. Therefore, we have that $\lim_{n\to\infty} f_n(1) = x^* > 0$, completing the proof of Claim 2. QED

$\square$

Ultimately, what does the theorem tell us? Informally, it's saying that we should worry about an infection becoming endemic if and only if a typical carrier infects more than one new healthy individual. Using our findings, we can extend our infection model to general networks, such as SIR models.

Definition. *A **SIR model** is a general disease infection model that defines the following disjoint subsets of the population of interest:*

- ***susceptible** individuals are people that can be infected;*

- ***infected** individuals are people that, for some period of time, are infected by a disease and can infect other susceptible individuals in the population;*

- ***removed** individuals are people that were previously infected, are currently not infected, and can never be infected again*